Grounded lexicon formation without explicit reference transfer

Vogt, P.

Published in:
Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)

Publication date:
2003

Citation for published version (APA):
Vogt, P. (2003). Grounded lexicon formation without explicit reference transfer: who's talking to who? In W. Banzhaf, T. Chistaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)* (pp. 545-552). Springer.

Grounded lexicon formation without explicit reference transfer: who's talking to who?

Paul Vogt

Induction of Linguistic Knowledge, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
Language Evolution and Computation Research Unit
University of Edinburgh, UK.
paulv@ling.ed.ac.uk

Abstract. This paper presents a first investigation regarding lexicon grounding and evolution under an iterated learning regime without an explicit transfer of reference. In the original iterated learning framework, a population contains adult speakers and learning hearers. In this paper I investigate the effects of allowing both adults and learners to take up the role of speakers and hearers with varying probabilities. The results indicate that when adults and learners can be selected as speakers and hearers, their lexicons become more similar but at the cost of reduced success in communication.

1 Introduction

In the past decade, an increasing number of alife researchers have investigated the evolution of language, see [3, 5] for recent overviews. One area of this field concentrates on the emergence of simple symbolic communication systems or *lexicons*. The focus of this paper will be on *lexicon grounding*, such as previously studied in [6, 7, 10]. This means that the evolved lexicons acquire their meanings through their interaction with their environment.

An example of a grounded experiment where agents develop a simple communication system is the *Talking Heads* [7]. In this experiment, a population of robotic agents developed a lexicon from scratch by interacting with each other by means of *guessing games*. In a guessing games a speaker attempts to produce an utterance to name an observed object, and then the hearer tries to guess the reference of this utterance. If the hearer fails to identify the correct reference, it receives corrective feedback from the speaker, which is used to learn the meaning of words. One problem with using *corrective feedback* is that it is not often observed in children's language acquisition [1].

In other approaches toward lexicon formation, the topic of communication was given to hearers prior to the verbal communication, see, e.g. [10, 12]. Recently an alternative language game has been developed independently by Smith [6] and Vogt [10]. In this approach, which I call *selfish games* (SG), agents do not give each other the reference of communication in a non-verbal manner, neither before nor after the verbal interaction. In [12] we have shown that a lexicon is learnable using the selfish game model, provided that language learners can learn from more experienced speakers. This result was shown using a simulation where we used iterated learning to model a population dynamics; if no population dynamics was used, the selfish game was only learnable for populations of 2 agents [6, 12]. The problem with this work was that the meanings were predefined and hence the simulations were subject to the symbol grounding problem [4].

The *Iterated Learning Model* (ILM) [2], has been designed to model a culturally transmitted evolution of language. In the ILM the population contains two groups of agents: adults and learners. The adults have passed the stage of learners and are supposed to have mastered the language, while the learners are novices who acquire the language by observing the linguistic behaviour of adults. After a given period of time the adults are replaced by the learners and novel learners are added to the population. In this model, adults take up the role of speakers in a language game, while learners play the role of hearers. In human societies, however, adults and learners both take the role of speaker and hearer. In the implementations of Brighton [2], this is not very important as the populations only contain one adult and one learner. However, when the population is larger, this has the effect that emerging languages may not converge, because the adults do not learn from each other, although they might invent new parts of the language [12].

In this paper, the selfish game and he ILM are combined to form a model of language evolution based on the idea that language evolved as a complex dynamical system [7] by means of cultural evolution [8]. I investigate some aspects of modelling language evolution and acquisition. In particular, I investigate the effects of both grounding and the way communication lines are distributed on a population's ability to form a shared lexicon. Although the model is intended to study aspects of human language evolution and acquisition, the purpose of this paper is to show the effect of these aspects on lexicon formation without concerning too much about its relations with actual human behaviour.

2 The method

The simulations presented in this paper were all carried out with the Talking Heads simulation tool *THSim v3.0* [11].[1] THSim can be used to simulate various aspects of language evolution and is based on the Talking Heads experiment [7].

As the (usage-based) selfish game model is explained elsewhere in this volume [11], this section only gives a brief summary. At the start of the selfish game, the agents observe a context containing 4 objects. In the current implementation, the agents only extract features relating to the RGB colours of the objects, so the lexicon they develop are colour names. The colours are generated with random values between 0 and 1 for each RGB component. Thus one might also say that the agents receive an arbitrary 3 dimensional feature vector.

The speaker selects one object as the topic of the game and the hearer has to guess what object the speaker selected. The speaker forms a meaning of the topic

[1] Downloadable from http://www.ling.ed.ac.uk/~paulv/thsim.html.

by categorising the extracted feature vector of the topic. Given a meaning, the speaker produces an utterance, which the hearer tries to interpret in relation to a meaning that refers to an object in the context. If the speaker fails to produce an utterance based on its lexicon, it expands its lexicon by inventing a new word. If the hearer does not know the utterance in relation to the meanings of the context objects, the utterance is associated with each meaning it does not know the word for yet. The speaker increases the usage of the used word-meaning association, the hearer increases the usage of each word-meaning association present in the context. The usage is used as a selection criterion for production and interpretation.

The selfish game is embedded in the iterated learning model, so after a given number of selfish games, the adults are replaced by the learners and new learners enter the population. This process then repeats for a given number of iterations (or generations).

3 Results

Given the models of the selfish game and iterated learning, a number of simulations were carried out, which will be presented in this section. The first simulation investigated the emergence of a lexicon in a population that used the SGs in combination with the ILM. In this simulation the speakers were all selected from the adult group and the hearers from the learner group, except during the first iteration where each agent is equally likely to be selected as speaker and hearer. This approach during the first iteration is taken, because it was found that otherwise no shared lexicon could emerge at all. From the second iteration onward the adults do not communicate among each other, nor do the learners. This is the control condition that reflects the original description of the ILM [2] and which was also applied in our previous (ungrounded) simulations [12].

In the second set of simulations, the probability with which the speakers and hearers were selected from the adult group was varied, thus allowing more lines in the communication. The probabilities are not selected to reflect any actual distribution of communication lines in human societies, although it should be clear that adults do not only communicate with learners as the ILM proposes. The purpose of this study is to investigate what effect the various probabilities have on lexicon formation.

In all simulations presented the population size N was set to 10. For each condition, 10 simulation runs were done for 10 iterations, where each iteration involved 50,000 SGs.

In order to analyse the effectiveness of the simulation two measures were used: *communicative success* and *similarity*. Communicative success measures, after each SG, the proportion of successful SGs over the past 100 games. The (lexical) similarity measure gives the proportion of agents $n(w)$ that use the same words w, weighted by the frequencies $U(w)$ with which the agents used the words, and averaged over the whole population of N. Similarity is calculated as follows:

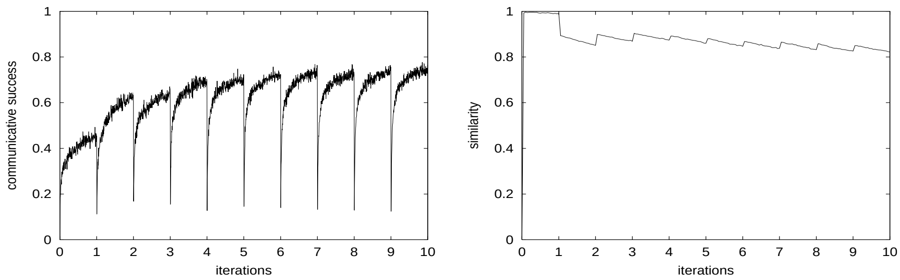

Fig. 1. The results of the simulations with only adult-learner interactions. The figures show (a) the communicative success and (b) the lexical similarity.

$$\text{sim} = \frac{1}{N}\sum_{w}\frac{n(w)\cdot U(w)}{\sum_{w}U(w)} \tag{1}$$

Similarity was calculated every 2500 games over the past 2500 games.

Adult-learner communication only. Figure 1 shows the results of the first simulation, averaged over 10 runs with different random seeds. Figure 1 (a) shows that the communicative success rose over iterations up to a value of around 75% in the final iteration and was still rising at the end of the iterations. It is noteworthy that the rise in success between the first and second iteration was much larger than between the other iterations. This has to do with the fact that, in the first iteration, all agents were equally likely to be selected as speaker or hearer. In between two successive iterations, the communicative success drops. This is because at those points new learners enter the population, and it takes a while before these new agents have acquired the lexicon from the new adults.

Although communicative success rose in subsequent iterations, this increase was small. More obvious is that the communicative success rose faster *within* succeeding iterations and was still rising at the end of the iterations. Because running all simulations of this paper is time consuming and running them for 50,000 games per iteration sufficiently illustrates the tendencies of the model, all simulations were run for 50,000 selfish games.

The lexical similarity revealed a different evolution, see Fig. 1 (b). This measure, which should ideally be 1.0, decreased from the second iteration onward to values slightly above 0.8. In addition, the trend within each iteration was decreasing. These findings reveal that the emerged internal lexicons (I-lexicons) of the agents was more similar in the first iteration than in following ones. This result should not be so surprising as from the second iteration onward the adults did no longer communicate with each other. As a consequence, the adults were not able to learn from each other and the learners learnt the different lexicons from these five different adults, causing the I-lexicons to diverge.

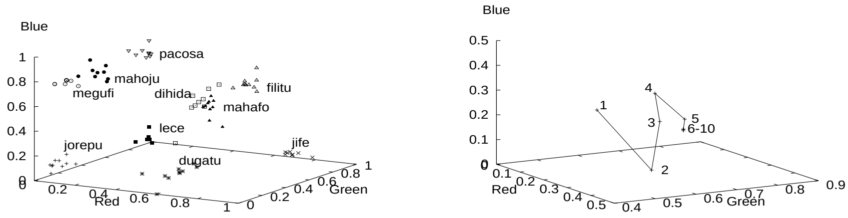

Fig. 2. Figure (a) shows a part of the lexicon in the RGB meaning space that emerged during the final iteration of one run of the simulation with only adult-learner interactions. Figure (b) shows the movement of the word "lece" over the different iterations. See the text for more details.

One of the main differences with the work presented here and that published in [12] is the lexicon grounding. Figure 2 shows some aspects of the emergent lexicons plotted in the meaning space. Fig. 2 (a) shows a part of the population's lexicons that emerged in the final iteration of one simulation run. The points in the RGB meaning space are the centre-of-mass prototypes that an agent used in association with a word in successful SGs. The prototypes shown belong to the 10 most frequently used words (the lexicons contained an average of about 110 elements). The plot shows that the prototypes used for a word are nicely clustered in the meaning space. This means that the distance between prototypes used in association with a particular word is relatively small. The plot also shows a more or less even distribution of the meaning space, which nicely corresponds to the distribution of the randomly generated colours.

Fig. 2 (b) shows the movement over the different iterations of the prototype for the word "lece" as used by the most dominant agent within an iteration. The movement starts in the first iteration around point $(0.4, 0.4, 0.3)$ and is $(0.1, 0.8, 0.1)$ from the sixth iteration onward. So after a relatively long walk through the meaning space, the meaning for "lece" becomes a rather stable point. This is because "lece" has become a word referring to the colours that are around this point, a property that is transmitted culturally to subsequent generations.

Inter-group interactions. In the second series of simulations the probability for selecting a speaker and hearer from the adult population (pS and pH) was varied. pS was varied between 0.5 and 1.0 and pH was varied between 0.0 and 0.5, both with intermediate steps of 0.1. The results of the simulations are shown in Figure 3.

The results are interesting. Fig. 3 (a) shows that when pS is 1.0 the communicative success is highest for all values of pH. When pS $= 0.9$, the communicative success is still fair when pH ≤ 0.2, but drops significantly for higher

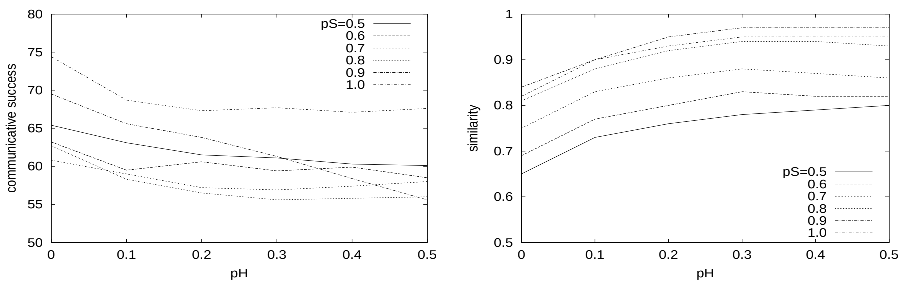

Fig. 3. The results of varying the inter-adult interaction probabilities. Figure (a) shows the communicative success and (b) the similarity both measured over the final 2500 games of the simulations. The y-axes show the different measures and the x-axes the different values for pH. The different lines in each graph relate to the different values of pS.

values of pH. The results are worst for pS $= \{0.6, 0.7, 0.8\}$, while the success for pS $= 0.5$ is higher than expected. It is also interesting to see that in all cases the communicative success more or less stabilised for pH ≥ 0.3, except for pS $= 0.9$.

The similarity reveals more structured results, see Fig. 3 (b). The similarity increases for higher values of pH, which is to some extent also true for increasing values of pS. The similarities for pS $= 0.9$ exceed the values obtained for pS $= 1.0$ for pH ≥ 0.2, but statistical tests revealed that this difference is not significant.

Extending each iteration to 100,000 instead of 50,000 SGs leads to an increase in communicative success of approximately 5%, while the similarity remained more or less the same. These results were obtained with four simulations for which $pS = \{0.5, 1.0\}$ and $pH = \{0.0, 0.5\}$.

4 Discussion

The simulation without inter-adult and inter-learner communication gives results that are, to a certain extent, similar to the results published in [12]. The main difference is that the level of communicative success is about 25% lower than in [12]. The reason for this lower performance is that, in the current simulation, the lexicon is grounded, whereas in [12] the meanings were predefined and shared by all agents. As a consequence, each agent develops its own ontology of meanings, which differ from each other. This, in turn, makes the learning problem in the grounded setting much harder. 75% communicative success is around the same level of success obtained for the observational games and guessing games in experiments using mobile robots [10]. Furthermore, the success is higher than the overall success of the Talking Heads experiment, which was on the average around 60% with large fluctuations [9]. This difference is partly explainable in the real world setting and partly in the uncontrolled population dynamics of the Talking Heads experiment. It is to be expected that – through the absence of corrective feedback in the SG – the Talking Heads experiment, which used guessing games, would outperform the current experiment [12]. Considering that the context size was fixed at 4, the results are significantly higher than chance, which is 25%. In addition, lexical similarity reveals that the lexicons are shared to a high degree. Furthermore, the results reveal an emergence of clustering of I-lexicons across agents. Hence, although the results are lower than those obtained in [12], the SG combined with iterated learning and grounding performs rather well.

In this paper the quality of the emerging I-lexicons was analysed using the newly designed similarity measure. The simulations show that similarity tends to decrease over different iterations when adults do not interact with each other. Starting from a shared lexicon in the first iteration – where all agents communicate with each other – the lexicons tend to diverge in following iterations. This differs from [12], where the opposite was observed. However, in [12] the distinction between adults and learners was made from the first iteration causing distinct I-lexicons between adults in the first iteration. Further research is required to explore if the opposing differences are indeed caused by the different conditions in the first iteration.

One interesting finding is that for fixed values of pS communicative success tends to decrease while similarity increases, which occurs when adults may be selected as hearers, i.e. when $pH > 0$. When adults communicate with each other, they learn from each other, causing their I-lexicons to become more similar. This similarity is in turn passed on to the learners of a population, which in the next iteration become adults themselves and pass on even more similar lexicons. The decrease in communicative success can be explained by considering two effects: First, when adults can be selected as hearers, the learners will receive less instances to learn from and consequently it takes more SGs to learn the lexicon. This is confirmed by the increase of communicative success in the extended simulations. The lack of change in similarity in these extended simulations further indicate that there is no causal relation between communicative success and similarity. Second, when learners are selected as speakers ($pS < 1.0$), they may attempt to invent parts of the lexicon themselves, which hampers the effectiveness of the communication, thus leading to a slower rise in communicative success.

5 Conclusions

In this paper the effects of grounding and varying communication lines on lexicon formation were studied using a simulation of the Talking Heads. In this simulation, the selfish game – that models communication without explicit reference transfer – and the ILM – that models cultural transmission over successive generations – were combined in a grounded setting.

The effect of grounding is that lexicon formation is harder than when the meanings are predefined; a result that is well known. In addition, the simulations show that the selfish game offers a method for constructing lexicons grounded from the population's interaction with the environment; a result that has not been shown before for populations larger than 2.

Varying the distribution with which adults and learners communicate among each other revealed that when speakers communicate more often with each other, the emerging lexicons become more similar. However, this is at the cost of a slower learning process among the learners.

Future investigations should concentrate on scaling up the system and perhaps allowing more realistic interaction strategies, which should reflect the psycholinguistic research.

Acknowledgements

The author is sponsored by a Visiting Research Fellowship awarded by the Royal Society of Edinburgh and the Caledonian Science Foundation. The author wishes to thank Andrew and Kenny Smith for the valuable comments made on earlier versions of this paper.

References

1. P. Bloom. *How Children Learn the Meanings of Words.* The MIT Press, Cambridge, MA. and London, UK., 2000.
2. H. Brighton. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54, 2002.
3. A. Cangelosi and D. Parisi, editors. *Simulating the Evolution of Language.* Springer, London, 2002.
4. S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
5. S. Kirby. Natural language from artificial life. *Artificial Life*, 8(3), 2002.
6. A. D. M. Smith. Establishing communication systems without explicit meaning transmission. In J. Kelemen and P. Sosík, editors, *Proceedings of the 6th European Conference on Artificial Life, ECAL 2001*, LNAI 2159, pages 381–390, Berlin Heidelberg, 2001. Springer-Verlag.
7. L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language*, Oxford, UK, 2002. Oxford University Press.
8. M. Tomasello. *The cultural origins of human cognition.* Harvard University Press, 1999.
9. J. Van Looveren. Robotic experiments on the emergence of a lexicon. In B. Kröse, M. de Rijke, G. Schreiber, and M. van Someren, editors, *Proceedings of the 13th Belgian/Netherlands Artificial Intelligence Conference, BNAIC'01*, 2001.
10. P. Vogt. Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, 4(1):89–118, 2000.
11. P. Vogt. THSim v3.2: The Talking Heads simulation tool. In *Proceedings of ECAL*, 2003. this volume.
12. P. Vogt and H. Coumans. Investigating social interaction strategies for bootstrapping lexicon development. *Journal for Artificial Societies and Social Simulation*, 6(1), 2003. http://jasss.soc.surrey.ac.uk.